



## The Application of Sentiment Analysis and Text Analytics to Customer Experience Reviews to Understand What Customers Are Really Saying

Gallagher, C., Furey, E., & Curran, K. (2019). The Application of Sentiment Analysis and Text Analytics to Customer Experience Reviews to Understand What Customers Are Really Saying. *International Journal of Data Warehousing and Mining (IJDWM)*, 15(4), 21-47. [2]. <https://doi.org/10.4018/IJDWM.2019100102>

[Link to publication record in Ulster University Research Portal](#)

### Published in:

International Journal of Data Warehousing and Mining (IJDWM)

### Publication Status:

Published (in print/issue): 01/10/2019

### DOI:

[10.4018/IJDWM.2019100102](https://doi.org/10.4018/IJDWM.2019100102)

### Document Version

Publisher's PDF, also known as Version of record

### General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [pure-support@ulster.ac.uk](mailto:pure-support@ulster.ac.uk).

# The Application of Sentiment Analysis and Text Analytics to Customer Experience Reviews to Understand What Customers Are Really Saying

Conor Gallagher, Letterkenny Institute of Technology, Donegal, Ireland

Eoghan Furey, Letterkenny Institute of Technology, Donegal, Ireland

Kevin Curran, Ulster University, Derry, UK

## ABSTRACT

In a world of ever-growing customer data, businesses are required to have a clear line of sight into what their customers think about the business, its products, people and how it treats them. Insight into these critical areas for a business will aid in the development of a robust customer experience strategy and in turn drive loyalty and recommendations to others by their customers. It is key for business to access and mine their customer data to drive a modern customer experience. This article investigates the use of a text mining approach to aid sentiment analysis in the pursuit of understanding what customers are saying about products, services and interactions with a business. This is commonly known as Voice of the Customer (VOC) data and it is key to unlocking customer sentiment. The authors analyse the relationship between unstructured customer sentiment in the form of verbatim feedback and structured data in the form of user review ratings or satisfaction ratings to explore the question of whether customers say what they really think when given the opportunity to provide free text feedback as opposed to how they rate a product on a scale of one to five. Using various Sentiment Analysis approaches, the authors assign a sentiment score to a piece of verbatim feedback and then categorise it as positive, negative, or neutral. Using this normalised sentiment score, they compare it to the corresponding rating score and investigate the potential business insights. The results obtained indicate that a business cannot rely solely on a standalone single metric as a source of truth regarding customer experience. There is a significant difference between the customer ratings score and the sentiment of their corresponding review of the product. The authors propose that it is imperative that a business supplements their customer feedback scores with a robust sentiment analysis strategy.

## KEYWORDS

Machine Learning, Sentiment Analysis, Text Analysis, Text Mining

DOI: 10.4018/IJDWM.2019100102

## 1. INTRODUCTION

Increasingly, leading modern businesses are looking to gain more insights from customer verbatim data they collect. Unfiltered customer feedback provides a tremendous opportunity to learn more about customer sentiment in relation to products and their end-to-end experience with a company. It also gives the business an opportunity to understand how they can 'close-the-loop' on any poor feedback they may receive and conversely how they can capitalise on positive feedback from customers. One of the drawbacks of online customer verbatim is its quality and quantity of data (Maritz, 2018). How do businesses mine and analyse data to ensure it reveals key insights that can ultimately drive profits in the right direction. Once a business recognises the benefit of analysing its customer feedback data, the question becomes, how? Identifying customer verbatim as unstructured data is the first step and secondly using a 'Big Data' approach such as text mining will allow the business to employ a more sophisticated and advanced modelling technique to uncover patterns in the data, employing sentiment analysis to identify key themes from the feedback. However, businesses will continue to use structured quantitative data gathering techniques such as asking customers to rate customer/client interactions, products and employees on various scales i.e. 1-5 Stars, 0 – 11-point scales and so on. These methods cannot be solely relied upon and it is key for a business to critically analyse its unstructured data whether than be on social media channels, ad-hoc emails to the business, letters and customer surveys.

All businesses need customers to prosper and grow. They are the main source of revenue for most businesses. It can be argued that the success of a business is directly proportional to its ability to acquire customers (Smith, 2016), keep customers happy, identify issues or irritants and consequently drive more selling opportunities. But for a business to achieve this it needs to identify the key indicators that will provide insights to facilitate this. There are many variables that a business needs to identify, the who, what, why, and how. Who are the potential customers in the marketplace? What do they want? Why do they want a product? How as a business can they retain customers and grow their market share. Examples of where a business can utilise their customer data are sales demographics, product and channel product preferences, social media or website sentiment and interactions and transaction behaviours. Customer analytics is becoming one of the key enablers available to companies to facilitate the translation of raw data into useful insights about their customers. This research intends to highlight the importance of customer analytics in the delivery of customer insights back to the business to drive decision making (Fiedler, et al., 2016). 60% of companies said that organisational silos were a major obstacle to improving customer experience (Google, 2016). Companies are finding it difficult to understand what customer data they have and often are underutilising the customer data that they do have (Department of Industry, 2018). The most successful companies have found ways around this and are actively implementing Customer Experience strategies i.e. building Customer Experience teams that are agile and can pivot based on business needs and goals (Hollyoake, 2009). While any company can use data to report financials or drive cost savings, the key differentiation comes from how this data can be used to drive business insights. While businesses are quite good at tracking structured feedback i.e. numerical ratings such as 1 - 5-star ratings they need to improve on understanding what customers are saying about their products and experiences with the company. This poses a serious challenge and difficulty increases as a business grows its customer base (Parasuraman, et al., 1991). With that in mind it seems impossible for a business to contact every potential and current customer individually to understand their sentiment. There are many methods or channels in which customers can interact with a business, for example, social media, email, customer satisfaction surveys, registering satisfaction via IVRs, capturing feedback on calls via Speech Analytics. So, the challenge for the business is how to accurately capture, store, analyse and obtain insights from this data. This becomes more difficult as the customer base grows, and it seems very unlikely that a business will create a touchpoint with every single customer they have.

In today's high-tech marketplace, customers have more options for products and services than ever before (International, 2017). The barriers of choice have come down as technology and the internet

has advanced and this leaves the business with a conundrum of how to accurately predict what their customers will do in the future before their competitors. To achieve this, they need to predict customer behaviour. This is where predictive analytics becomes the answer. To get ahead of the curve and ultimately their competitors (LaValle et al., 2011) argue that a business needs to proactively employ advanced data analytics techniques and drive a strategy that will provide insight into what their customers might do in the future. What products are customers buying? What is the likelihood that a customer will churn and move to a competitor? Are customers satisfied with the products, if not why? Are there irritants that customers find when dealing with the business? There are a multitude of questions the business will have, and predictive analytics can go some way in providing insights. Using advanced data analytics techniques, a business can use their customer data to build predictive models. These models have the potential to predict future customer behaviour. It aids businesses to identify customers at risk of leaving, potential issues with products or services. With predictive analytics, businesses can gain a competitive advantage over rivals if used properly. There is a trend towards prescriptive analytics and a move away from descriptive analytics (S. Kaisler, 2013). This allows the business to focus on what could happen, based on predictive modelling methods rather than what has already happened. There is obviously a requirement for descriptive analytics, businesses will always have to report what has happened at a certain point of time hence the need to capture metrics and KPIs but the notion of utilizing this data to predict what will happen shows a forward-thinking mindset (Steger, 2013). The opportunity for exploring predictive customer analytics today is better than ever before. Data sources such as customer satisfaction surveys, social media channels, chatbots, voice calls using speech analytics all provide a rich source of opportunity. And naturally Big Data technology has kept pace with the exponential growth in data and is capable of handling large-scale data processing, storage and integration needed to capitalize on these uncovered insights (EMC, n.d.) To begin to understand what the customers think and feel a business needs to understand the type of data it has and build the technological infrastructure to be able to mine the data to build effective models (McKinsey, 2016). Only at this point can the business develop these models effectively and drive their analytics strategy forward

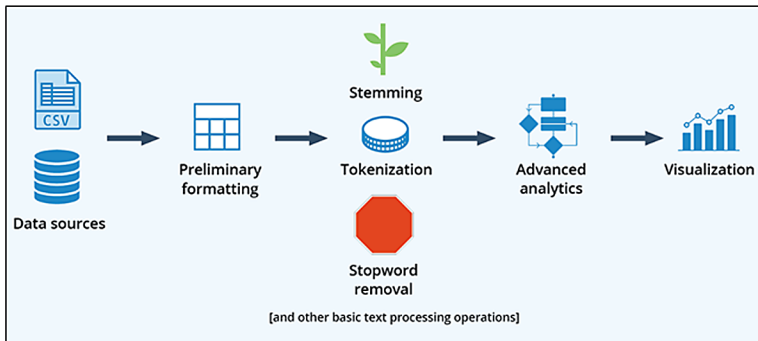
We examine if it is possible for a business to rely solely on a single point Customer Satisfaction metric through the value of sentiment contained within customer feedback as an indicator of customer satisfaction and/or advocacy with a company or product. Essentially, we investigate if a customer sentiment provides a truer reflection of the customers experience with a service or product.

## 2. RELATED WORK

Text analytics provides a set of techniques to make trends visible, dynamic, and to reveal relationships between data points that may have been difficult to uncover by a human due to the scale and quantity of the data (Miner, 2012). Numeric ratings remain the most generic form of data collection in customer surveys and unstructured feedback such as customer verbatim continues to increase in importance. Rather than attempting to manually review reams of customer comments, tools like text analytics and speech analytics allow companies to turn this verbatim feedback into valuable insights they can analyse and measure more efficiently than manually reviewing the data (Verint, 2016). The end to end Text mining process has many stages all of which are key to the end goal of deriving insights from the textual data. Figure 1 gives an idea of how this would look as a process flow.

It is key that the data needs to be in a processable format. Multiple data sources can be used, a large amount of textual data may be available via free text format on social media platforms, customer surveys and speech analytics tools would also convert conversations into a raw format. Preliminary formatting is preparing the data by removing dirty or erroneous data. This improves the data and lessen the chance of errors although can be time consuming (Natarajan K., 2010). Other formatting procedures can be applied to the data as required. Textual data contains lots of small yet highly frequently-used words: “at,” “the,” “or,” “for” etc. To analyse the frequency of words in a text block,

Figure 1. End to End Text Mining Process (Harris, 2018)



these types of words will always occur. Stop word removal is the process of filtering out these words to look for longer, more significant words. Python for example comes with built-in stop word dictionaries for different languages (Upadhyay, 2017). “Tokenization looks at the spaces between the words as word boundaries to ‘cut’ words from the text string at those points.” (Ott, 2018). Tokenization is a crucial step in the process of structuring text for further analysis, as otherwise text is treated as an unstructured “string.” The idea is to break the string into individual words or tokenize, the sentence or piece of text. This allows for the determination of word frequency in the document and begin identifying relationships between words. The intention for a business for example would be to ensure that frequently-used words are tokenized properly with a view to further analysis. Most of text-mining tools allow users to create customized tokenization rules. The Bag-of-words process, also known as a vector space model, breaks down a document into its constituent words, regardless of grammar and order. This is a crucial step as there is the likelihood of receiving additional textual data which require be broken down in similar fashion to allow for further analysis (Robert P. Schumaker, 2012). Stemming (a.k.a. “lemmatization”) is the process by which words are reduced in a document to their basic form. Taking the English language as an example, words have many forms rather than one single form based on their grammatical purpose. For instance, “was,” “are,” “is,” etc., are all forms of the verb “to be.” As with Tokenization above, once you perform a process like stemming then the data is prepared for further analytics which can be applied to textual data to recognize patterns (Mejova, 2012). One example of an advanced analytics technique associated with the text mining process is Cluster analysis. In the context of the text mining process, cluster analysis involves classifying words into groups, such that words in each group are more related to one another than they are to words in other groups. (Kochut, et al., 2017) This is done through the application of algorithms such as k-means and x-means clustering. Using K-means clustering the variable ‘K’ represents the number of clusters you want. You could potentially segment data into four groups, eight groups — then the algorithm understands that and executes based off that ‘K’ number. For x-means the number of optimal clusters is unknown and here you would specify a max and min number of clusters. The algorithm will then run an analysis and output the optimal number of clusters. Clustering is a powerful way to identify relationships and trends in copious amounts of textual data (Allahyari, 2017). It can also be used as part of the sentiment analysis process. Data Visualization plays a key role in allowing a business or researcher to identify key trends or insights from the data. There are multiple methods to visualise data, one could use the inbuilt features of a programming language e.g. Python has matplotlib for creating charts and visuals from data. There are a vast range of Data Visualisation software available from a Business intelligence point of view, i.e. Tableau is an industry leader in Data Visualisation (Ajenstat, 2017), Microsoft’s offering of PowerBI is an up and coming challenger to Tableau (Molag, 2017). These business intelligence tools can link to multiple data sources i.e. excel, csv, SQL server, json and interestingly they now can link directly to Python or R scripts, so that users can run scripts

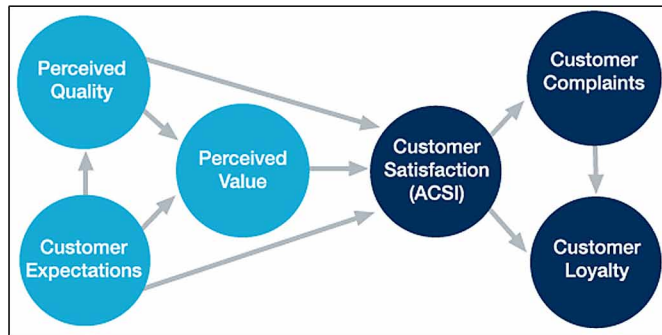
inside the tool and then create visuals and ultimately dashboards to aid in the insight exploration of the data. This is extremely beneficial in a business scenario where companies can have dedicated Data Visualization analysts who create and maintain these types of dashboards. As part of this research the author plans to utilise and promote the benefits of visualising the data so that the reader can understand how this would play out in a business environment.

Text mining can be utilized to evaluate customer satisfaction from verbatim feedback rather than relying solely on a numerical indicator from a customer satisfaction survey-based (Hosseini, et al., 2010) proposed a data-mining model using monetary (RFM) attributes and K-means clustering to evaluate customer loyalty and customer lifetime value (CLV). The research investigated the degree to which customer loyalty influenced bottom line profits for an organization. The model developed showed a clear improvement in the accuracy of the customer loyalty measurement. This study had limitations as the model relied too heavily on combining data-mining techniques alone and did not consider the impact of textual data such as verbatim comments from customers, which are now generated at many touchpoints in the customer's journey. (Pang & Lee, 2002) employed text mining methods to extract insights from customer data using sentiment analysis to try to improve customer loyalty measurement. Another example of using text analytics to understand the dimensions of product alignment to gain insight into brand positioning was used by (Tirunillai & Tellis, 2014). (Ordenes, et al., 2014) found that text analytics could be used to analyse customer feedback to capture insights of the customer experience. This model captured customer sentiment from the data and tagged it as either complaints, compliments or suggestions/feedback. This allowed them to propose insights to suggest efficiencies for the company and its resources. Their research, however did not combine the qualitative data and quantitative data to assess and predict customer loyalty. Where the research proved useful was in relation to the computation methods applied that could produce findings that can be expressed quantitatively, and this for example aids the desired outcome to conduct statistical modelling, data visualization, and machine learning algorithms to derive further insights.

Businesses have two kinds of customer feedback data that they store, analyse and measure: structured and unstructured data. Structured data is information that is clearly defined and easy to report on. It is the kind of data that is generally found in a survey and can be organized in a spreadsheet: name, location, age, and rating (3 out of 5 stars, for example, or a 10 for "most satisfied" versus a 1 for "least satisfied") (Taylor, 2018). Unstructured data as it exists today in the context of Customer experience data is, basically, text or free form customer verbatim, although it can also include other media such as audio, photos, or videos. Unstructured data can be captured in an email, the "additional comments" section of a survey, voice recordings of customer interactions e.g. Speech Analytics tools, a post on a customer review site e.g. Amazon, in social media, and in a multitude of other places. (Taylor, 2018). Analysing all this data correctly is critical, because it reveals everything from buying trends to product flaws and provides a significant business advantage. (Aswani, 2017) Organizations often struggle to do this analysis, however, because unstructured data is significantly harder to categorize and report on than structured data. It can be hard to parse due to grammatical errors or slang, it frequently contains multiple unrelated ideas, and it can represent various levels of sentiment related to each idea (e.g. "I couldn't complete the registration process online, I didn't understand the website, but the rep was very patient and talked me through the process."). Unstructured data is, by its natural harder to navigate than structured data. Anecdotally it is thought that 95% of customer feedback data is unstructured (Kemper, 2008) and it contains a wealth of information to aid in the understanding of customer sentiment and opinion. Daunting as it may seem to businesses, this type of feedback is predicted to continue to grow, as millions of people increase their use of online (Clarabridge, 2015). Customer satisfaction is one of the most critical issues concerning business organization of all types, which is justified by the customer-oriented philosophy and the principles of continuous improvement in modern enterprise (Arokiasamy, 2013) see Figure 2.

Net Promoter Score (NPS) is a measure of customer loyalty and is often held as the leading customer experience metric. NPS is still a popular customer loyalty measurement despite recent

Figure 2. Basic American Customer Satisfaction Index model (ACSI, 2018)



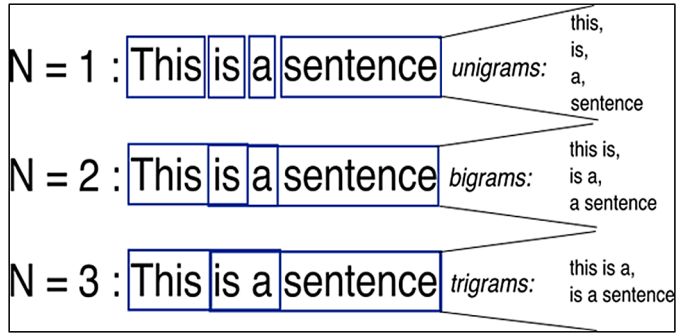
studies arguing that customer loyalty is multidimensional. Therefore, firms require new data-driven methods that combine behavioural and attitudinal data sources. Measured on a 0 -10 scale, it is broken down where a score of 0 – 6 is classified as Detractors, unhappy customers how are very unlikely to recommend the business to friends or family. A score of 7 – 8 is classified as Passives, those customers who are happy enough with the service but not enough to actively encourage others and a score of 9 – 10 is classified as Promoters, those customers who are loyal and would recommend the business or product to friends and family (Qualtrics, 2018).

Sentiment analysis plays a key role in the development of predictive and machine learning models. It has developed into a wider all-encompassing process that involves analysing textual data and applying a ‘sentiment score’ to that text (Mejova, 2012) One of the most widely used approaches is using machine learning strategies to develop and build an algorithm using a training data set before applying a test or ‘real’ data set. Machine learning techniques first trains the algorithm with some inputs with known outputs so that later it can work with new unknown data (Devika, et al., 2016). SVM or Support Vector Machines are defined as a supervised machine learning algorithm normally used for regression and classification processes. SVM uses a non-probabilistic classifier in which a large amount of training set is required. It is done by classifying points and identifying a (d-1)-dimensional hyper plane. SVM looks for the hyper plane that best segregates the classes (Devika, et al., 2016). Support Vector Machines make use of the concept of decision planes that define decision boundaries. A decision plane is one that best differentiates between a set of objects having different class membership. Support Vector Machines are considered a powerful classification algorithm. When used in conjunction with other techniques such as random forest and other machine learning tools, they give a very different aspect to ensemble models (Srivasta, 2014). There is an argument that SVM do not perform well with larger data sets due to the high training time required, they do however perform well with clear margin of separations between objects (Brownlee, 2016).

In the context of sentiment analysis or text mining, an n-gram is a neighbouring sequence of n items from a given sequence of text or speech. N Items can be letters, words, pairs of words for example and are typically collected from a sequence of text or speech corpus. Figure 3 shows how a sentence can be broken up into various formats i.e. words, pairs of words, etc. N-gram sentiment analysis considers the sentence as a whole (Ghiassi et al., 2013).

In Machine learning, Naïve Bayes is termed as a family of “probabilistic classifiers” based on the use of Bayes Theorem. Naïve Bayes classifiers are highly scalable and are mainly used when the size of the training set is less. The conditional probability that an event X occurs given the evidence Y is determined by Bayes rule shown in Equation (1). Unlike Naïve Bayes, Maximum Entropy Classifier does not assume that the features are conditionally independent of each other. Uncertainty is maximum for a uniform distribution and the measure of uncertainty is known as entropy. Maximum Entropy Classifier can be used to solve a variety of text mining scenarios particularly sentiment analysis

Figure 3. N-Gram example (Maiolo, 2015)



(Vryniotis, 2013). Maximum Entropy Classifier uses a parameterized set of weights which when combined, join the features that are generated from a set of features by an encoding. This encoding then maps each pair of feature set and labels them to a vector (McCallum & Nigam, 1998). ME classifiers are known as the exponential or log-linear classifiers and because they work by extracting sets of features from the input data, combining them in a linear fashion and then using this sum as the exponent (Daume III & Marcu, 2006). (Pang & Lee, 2002) used the approach that each document is represented with an array of 1's and 0's that indicate if a word exists in the text of the document. This plays on the BoW (Bag of Words) theory commonly used in NLP and text mining.

$$P(c | x) = \frac{P(x | c) P(c)}{P(x)}$$

$\downarrow$  Likelihood       $\downarrow$  Class Prior Probability  
 $\uparrow$  Posterior Probability       $\uparrow$  Predictor Prior Probability

$$P(c | x) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c) \quad (1)$$

The K-Nearest Neighbour (KNN) technique is based on the notion that the classification of an instance will be in some way like those nearby it in the vector space hence the term nearest neighbour. (Srivastav & Singh, 2014) researched weighted k-Nearest Neighbour where they assumed weightage to those objects in the training set and then used these weights for their calculation of sentiment of text in word by word fashion. In a weighted k-NN analysis (Srivastav & Singh, 2014) the sentences of the tweets captured were tokenised and then stop words were removed. The algorithm developed a positive score to each review after the first parse. For the second parsing process a neutral review is input. This score is then modified if required allowing for better positive determination and creating an output file that consists of the review ID and its associated positive score. Weighted combinations of feature sets can be quite effective in the task of sentiment classification, since the weights of the ensemble represent the relevance of the different feature sets (e.g. n-grams, POS, etc.) to sentiment classification, instead of assigning relevance to each feature individually (Xia et al., 2011). (Balahur & Montoyo, 2008) investigated the notion of assigning products classes to textual data sets and extracting general features from the products and then in turn assigning sentiment to the features.

A Rule based approach (Liu, et al., 2005) to extract opinion from a textual data set by defining various rules tokenized each sentence in every document of the data set and tested for the presence of each token or word. If the presence of the word is confirmed and has a positive sentiment, then a



“+1” rating is applied to it. Each piece of text starts with a neutral score of zero and was considered positive. In (Devika, et al., 2016) for example, if the input sentence contains any word which is not present in the database then it will be added to the database. This is a supervised learning approach in which the system is trained to learn if any new input is given. Rule-based approaches tend to use heuristics to determine sentiments and linguistics research to analyse sentiments. They tend to have very poor generalization, but within a narrow domain they tend to perform well (Medhat et al., 2014). In machine learning this is known as overfitting and the aim is to avoid this problem and conversely with rules-based systems it's considered largely unavoidable. (Devika et al., 2016) employed a machine learning data driven approach which used a labelled corpus of texts with their respective sentiments for prediction.

A Lexicon based approach uses dictionaries of words tagged with their semantic polarity and sentiment strength (Yadav & Elchuri, 2013) These scores are then used to calculate the polarity and/or sentiment of the textual data. It is said that this technique allows for high precision but low recall. Lexicon Based techniques work on an inference that the aggregated semantic polarity of a sentence or tokens is the sum of polarities of the individual phrases or words. At the ROMIP 2012 conference the lexicon-based method proposed in (Chetviorkin & Loukachevitch, 2013) was used. The authors employed an emotional research technique for sentiment analysis dictionaries for each of the domains. Using appraised words from the appropriate training set, the domain dictionaries were reloaded with the highest weighted words

### **3. METHODOLOGY**

We built a classification model using predictive variables. The steps were:

1. Apply a linguistic text mining approach (Ordenes et al., 2014) to determine the category of each customers review by mining the review verbatim thus dividing customer reviews into three groups of positive, neutral or negative. This sentiment analysis will facilitate the process of modelling the sentiment score generated against the customer rating score, hence attempting to predict the likelihood that the customer has rated the product 1 thru 5.
2. Apply a Naive Bayes Machine Learning Algorithm to predict the rating of the reviews as this is found to work well with the text data.
3. Prepare the textual data for vectorization by removing the punctuations and then converting into lower case and all stop words removed from the sentences.
4. Convert Review' text data to a token (which has punctuations and stop words removed). Using Scikit-learn's Count-Vectorizer functionality the text is then converted into a matrix of token counts. It effectively is a 2D matrix where each row is a unique word, and each column is a review.
5. Develop the prediction model using a Naïve Bayes algorithm to predict customer satisfaction scores accuracy.
6. Evaluate the model by testing the predicted values against the actual ratings in the data using a confusion matrix and classification report from the Scikit-learn toolkit. The data is run through the model and the efficiency of the model is noted and commented upon.
7. Test the output from the model to confirm that predictions are accurate and line up with the expected results. This involves testing random reviews and noting the predicted outcome from the model and comparing with the actual rating in the data.

#### **3.1. Data Transformation**

The data set was sourced from Kaggle<sup>1</sup>. It comprised mobile phone reviews from the Amazon website (n=413,840) with only 62 not having a verbatim comment in the 'Review' column. The key component

we looked for was the rating (start rating or 1-5 rating for example) and a corresponding review from the customer. We wanted a wide-ranging variance in the types of reviews left by the customers so that there is some degree of subjectivity as differing views are important to aid the design and development of the sentiment model. Therefore, there is an expectation to see different results from the sentiment analysis of this dataset compared to that of datasets about technology products. The variables were Product Title, Brand, Price, Rating & Review text and number of people who found the review helpful. Python was used as it has a considerable variety of text mining and sentiment analysis packages (Bird, et al., 2009) and packages that allow the visualisation of the data and it has a solid range of predictive analytics packages.

Transforming customer satisfaction data into customer insights requires a formal system by which measures are included as part of a data collection and analytics process. The aim was to investigate the correlation between the sentiment analysis score and that of the customer rating score and predict with a degree of certainty the likelihood of the rating based of the sentiment? Therefore, rather than relying on the survey-based CSAT measurement, we applied big data techniques to the mobile phone customer data to evaluate the link between satisfaction and sentiment. The basis of this analysis is to study the customer feedback from the 'Review' column in the data containing the textual feedback, to understand the sentiment behind their commentary and to identify key themes and words within this data. To do this there are several stages required and this follows the Cross-Industry Process for Data Mining (CRISP-DM) (Chapman, et al., 2000). The CRISP-DM process is the "de facto standard for developing data mining and knowledge discovery projects" (Marbán et al., 2009). This multi-layered research methodology has 5 key components of which this process utilizes some but not all the elements of the framework: (Marbán et al., 2009).

Customers who purchased a phone from Amazon were invited to rate the phone i.e. Customers Satisfaction score or 'Rating' as per the data, provide a review of the product via a free form text box i.e. 'Reviews' and as per the amazon website other users could react to the review and 'up-vote' the feedback given by the customer on the product. The customer experience strategy of capturing data like this has been touched upon previously but it is relevant to understand how it is applicable in a business context. As a business there must be some understating of how customer react to products and in this data set there are two key variables that provide insight into this i.e. Rating and Reviews. We note that this is an amazon data set and not company specific data i.e. Apple do not own this data for example and whilst they could carry out analysis on the data do not have sole ownership of it.

The data set can be considered longitudinal customer data covering attitudinal data elicited from the customer survey, which includes Satisfaction ratings ('Rating') and qualitative customer verbatim comments ('Reviews'), behavioural data (transactional data i.e. price) and product data across multiple products for all major mobile phone companies. The data source did not have any date/time references associated with the records. Therefore, the opportunity to conduct time-series analysis is unavailable, this would be an area where the author deems future work would be recommended. There was a total of 413,840 product review records.

The key predictive variables in the data sets that are considered for analysis were, Rating and Review. It should be noted that there will be descriptive analysis performed on the all the other variables within the data set to look for additional insights from the data. The intention is to allow the business to accurately predict the likelihood that a customer will rate a product the way they have. The process will involve identifying those customers who rated the products 1 thru 5 and then using a predictive modelling technique output the accuracy using a classification report.

### 3.2. Analysis

Pearson and Spearman correlation plots ascertain where there are correlations both positive and negative in the data and this will then help decision making easier when deciding which variables to look at as part of the descriptive analysis in later sections. Data correlations help to identify related variables and can be used by investigating the relationship between two quantitative, continuous

variables. Pearson's correlation coefficient ( $r$ ) is a measure of the strength of the association between the two variables and the interval level ranges between -1 to +1. We use Spearman's correlation which is the nonparametric version of the Pearson correlation. Spearman's correlation coefficient, ( $\rho$ , also signified by  $r_s$ ) measures the strength and direction of association between two ranked variables (Hauke & Kossowski, 2011).

The first part of the analysis was done using the matplotlib library and a comparison was run between the Review and Review length variables to get an understanding if there was any significant correlation between the length of a review given by the customer and the subsequent rating. Histograms were generated representing each review category i.e. 1 -> 5 showing the distribution of review lengths for each rating. A comparison was carried out to analyse how the rating and price affected each other. Using the Seaborn library, a scatter graph with a regression line was created for each rating in relation to the corresponding price. For this analysis the heatmap functionality from the seaborn library was used and this produced a heat map graphic that grouped the numerical variables in the data (Price, Review Votes and Review Length) and assigned a colour and correlation score to each relationship indicating the strength of the correlation. This is useful in the sense that it allows the reader to spot any important positive or negative correlations between the data points that may not have been obvious when eyeballing the raw data. Based on the results one could then decide on how best to continue analysing the data.

A scatter chart from the seaborn library was used in this analysis to look at the distribution of review lengths across each rating. In this scenario the data is plotted using a scatter plot and the reason for this is to identify if there are any outliers in the data across the 5 ratings. To investigate the Ratings distribution of the data a KDE (Kernel Density Estimation) plot was used to visualise the ratings. In statistics, KDE is a non-parametric way to estimate the probability density function of a random variable, in this case the Rating given by customers. KDE attempts to smooth the data where assumptions about the population are made based on a finite data set as shown below in Equation (2).

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (2)$$

where  $K$  is the kernel, a non-negative function that integrates to one and  $h > 0$  is a smoothing parameter called the bandwidth. In Equation (2), the kernel  $K$  with subscript  $h$  is called the scaled kernel. Ideally one would look to select the  $h$  value as small as is feasible for the data; however, there is always a trade-off between the bias of the estimator and its variance. The bandwidth ( $h$ ) controls how tightly the estimation is fitted to the data, very much like the bin size in a histogram. The KDE plot is very similar to a histogram, it estimates the probability density of a continuous variable, in this case the probability of a rating from 1 to 5. In this case a univariate or kernel density estimate is plotted. A comparison between Apple and Samsung was plotted in two separate graphs allowing for the analysis of the distribution of ratings between the two brands. A side by side comparison can be valuable in determining any obvious differences between the brands with future analysis in mind. Staying with KDE plots and using Samsung again as an example, a plot was created to visualise the distribution by Price of the Samsung products using the Seaborn library as in the previous two sub sections.

### 3.3. Sentiment Analysis

Using the TextBlob library, we conducted sentiment analysis on each comment for each dataset using TextBlob and assigned each 'Review' a polarity score of between -1 -> 1. The ranges for Positive, Negative and Neutral were Positive Sentiment Score: >0.1 to 1, Neutral Sentiment Score: between -0.1 and 0.1 and Negative Sentiment Score: -1 to <-0.1. These ranges represent the various categories of sentiment polarity. A data frame named 'df\_polarity\_desc' was used to output a table

with the columns Review, Sentiment and Polarity. This enabled the reader to view the output from the sentiment analysis and to put some context around the actual Sentiment Polarity scores for each individual review. Two separate word clouds were created to visual both 'Positive Words' and 'Negative Words'. Both 'Positive Reviews', 'Neutral Reviews' and 'Negative Reviews' were defined based on the numerical ranges discussed. This allowed the author to categorise the Reviews into the three 'buckets' described.

Using the Word Cloud function, two words clouds were visualised for both Positive and Negative reviews i.e. the most used words in each categorisation. A word cloud with a black background was produced for the Positive Review words and a word cloud with a white background for the negative review words. The Sentiment Polarity score was generated on a -1 to 1 scale based on the corresponding 'Review'. The intention is to test whether the sentiment polarity score is related to the rating provided for the corresponding text comment from the customer. To achieve this the Sentiment score has been normalised from a -1 to 1 scale to a 1 to 5 scale to mimic the ratings scale available to the customer. To normalize the data, Equation (3) was used and applied to the 'Polarity' score and the normalised data was named 'Polarity\_Norm'.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

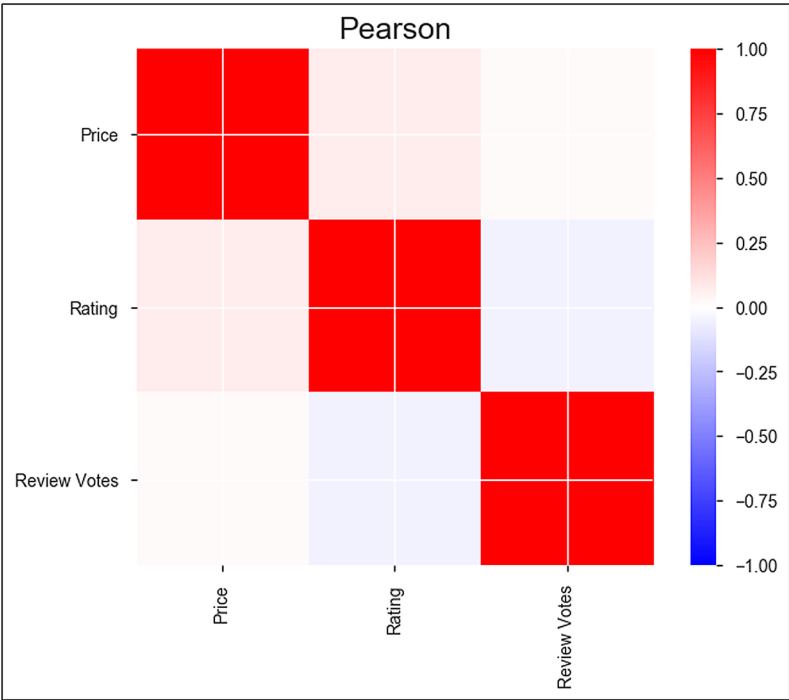
The Polarity scores were normalized as per the above to scores between 1 and 5 to allow for direct comparison to the corresponding rating score. Once the Polarity score has been normalised, the mean and standard deviation of 'Polarity\_Norm' is used to create several charts analysing the probability distribution for each Rating category. The two main rating categories that have been identified throughout the research have been Rating 1 and Rating 5. A plot of the polarity distribution is created to explore the sentiment from the reviews in terms of the actual rating and this provides a key insight into the data.

Word frequencies are gathered from the Reviews textual data set into a document-term matrix thus allowing the extraction of features from the Reviews sub dataset. We can employ a feature extraction technique that allows for transforming arbitrary data, in this case the Review text, into numerical features usable for machine learning. The vectorization process turns the collection of text documents into numerical feature vectors. This technique employs tokenization, counting and normalization of the data and is commonly referred to Bag of Words or "Bag of n-grams" representation. Using 'bow\_transformer.vocabulary' a classification technique is carried out on the data. Since the vectorization process provided discrete features i.e. the word counts from the Review data the next step is to employ Multinomial Naïve Bayes with the feature counts (Huang, 2017). This helps us calculate the probability of two ratings categories, Rating 1 and Rating 5, using Bayes theorem. Using Reviews as the second variable in the testing, the probability and accuracy of predicting a Rating of 1 and 5 based on the reviews will be obtained. To do this the data is split into a testing and training set. We take 30% of the data and use that as the testing set. Using a confusion matrix, we summarise the performance of the classification algorithm. The number of correct and incorrect predictions are summarized with count values and broken down by each class. We run the confusion matrix for 1 and 5 Ratings and then all Ratings i.e. 1 thru 5. Due to the nature of the confusion matrix it is expected that the accuracy will drop when the 5 classes are modelled.

#### 4. EVALUATION

The Pearson Correlation showed a slight correlation between Rating and Review Votes (see Figure 4). This would suggest that as the Rating increases the price increases and the number of Review votes

Figure 4. Pearson Correlation



given to the product increases also. Which would make sense but since there is not a strong correlation the thought is that this would be not worth exploring with further analysis. Using the Spearman correlation in Figure 5, one can observe very similar correlations with a slight negative correlation between Rating and Review votes which, as with the Pearson test, does not warrant deeper analysis.

The correlation between variables showed a negative correlation between Price and Polarity which would suggest that the sentiment of the review tends to be more negative as the price goes up. This potentially could indicate that customers are less happy with their more expensive phones and expect better from the product. Polarity and Review votes have a positive correlation with each other, indicating that the more a Review is 'up-voted' the higher the more positive the sentiment seems to be, which would make sense from a logical standpoint (see Figure 6).

We created histograms to compare review lengths against each rating to help us note the number of reviews and which buckets the review length belong to for each rating. This reveals that there is an overwhelming number of reviews for Rating 5 and a considerable number of reviews for Rating 1. This will be useful going forward in determining which Ratings should be more closely looked into. Figure 7 is quite interesting as it displays the distribution of price points across each Rating value. As expected, many price points exist below the \$1000 range with a small number of outliers upwards of \$2000.

Using a heat map to look at correlations between the variables it is noticeable that there are negative correlations of -0.91 between Price and review length which would indicate that the greater the price the shorter the review and vice versa. This could be attributed to the fact that there are more reviews at the lower price level as noted previously in Figure 8 and therefore this finding it not at all surprising. Like other charts, it was felt that a closer look should be taken at Review length v Rating. (see Figure 9). What is observed here is the distribution of reviews by Length v Rating and it shows that there is a slight tendency for customers to comment more about phones that they are happy with hence the distribution in Figure 10.

Figure 5. Spearman Correlation

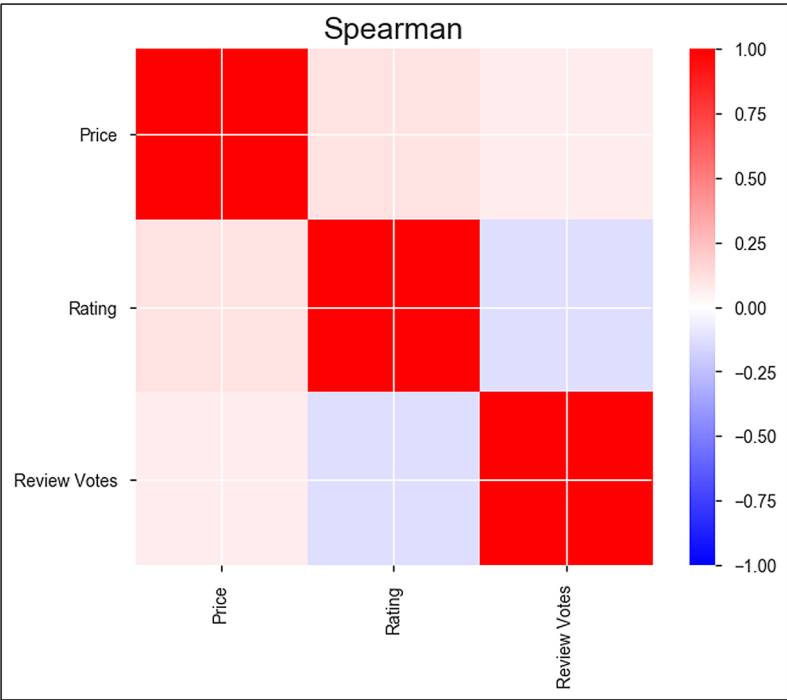
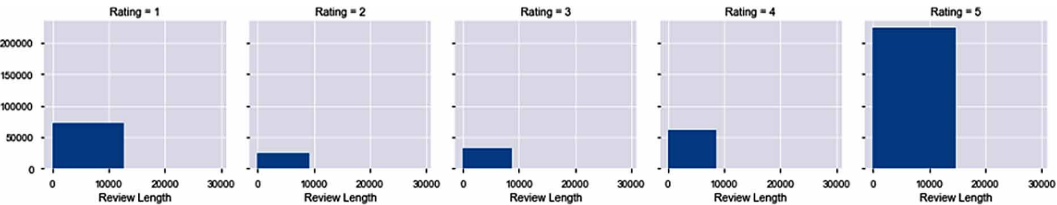


Figure 6. Review v Review Length by Ratings



To get a sense for the multivariate density of the ratings the chart in Figure 11 shows the high density of % ratings across the reviews. There is a smaller but significant population of 1 Star reviews and this will supplement the previous charts for further investigation. Comparing Figure 11 with Figure 12 there is no distinct difference between the two distributions indicating that there is no significant difference between brands.

4.1. Word Frequency

The word phone is by far the most common. Some other interesting outputs from this are the occurrences of the words Great and Good which would lead to a possible conclusion that there may be an overwhelming number of positive reviews in the data. Figure 13 shows the Top 7-words.

As another visual alternative Figure 14 shows a word cloud of word frequencies. This gives an impactful visual of the occurrences of the top 100 words in the review feedback.

Using the TextBlob library the intention is to give each ‘Review’ a sentiment score of between -1 -> 1. These define the various categories of sentiment. There is an overwhelming number of positive

Figure 7. Regression Plot of Price v Rating

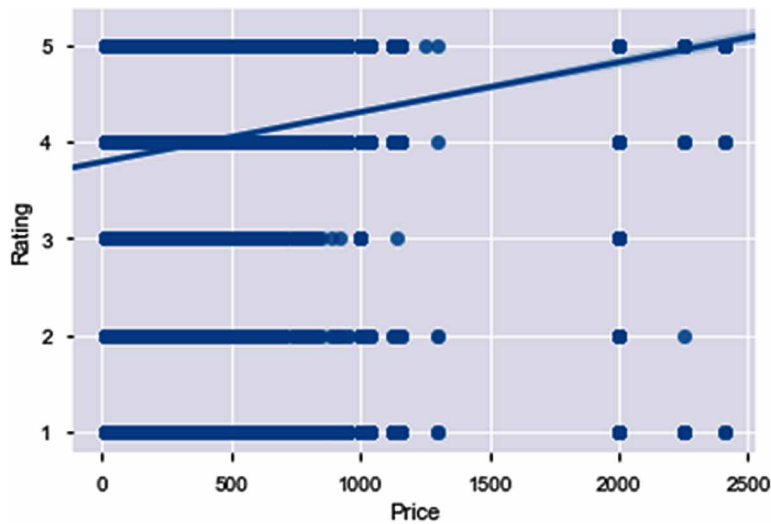
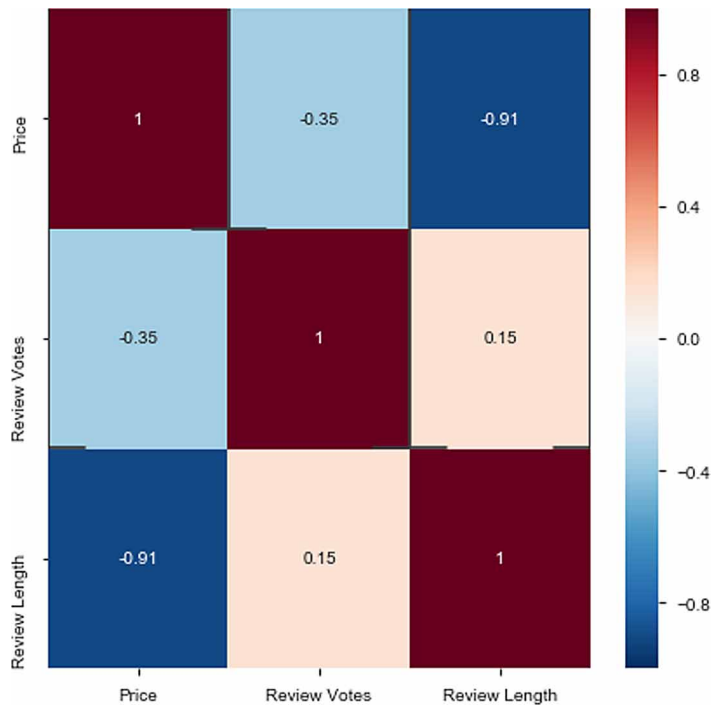


Figure 8. Heat Map



reviews and this follows from the exploratory analysis above showing the high frequency of 5-star ratings. This does not tell the whole story so further analysis on the sentiment is required.

The results in Figure 15 show that sentiment analysis results mostly fall into the Positive Review category based on the range used. Neutral Reviews seems to have a reasonable volume with the narrow range taken into consideration and there are a smaller number of Negative reviews compared

Figure 9. Rating v Review Length Regression plot

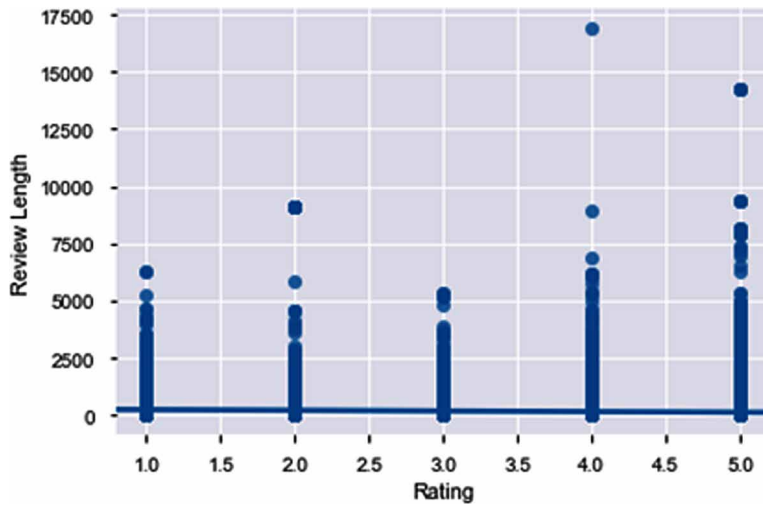
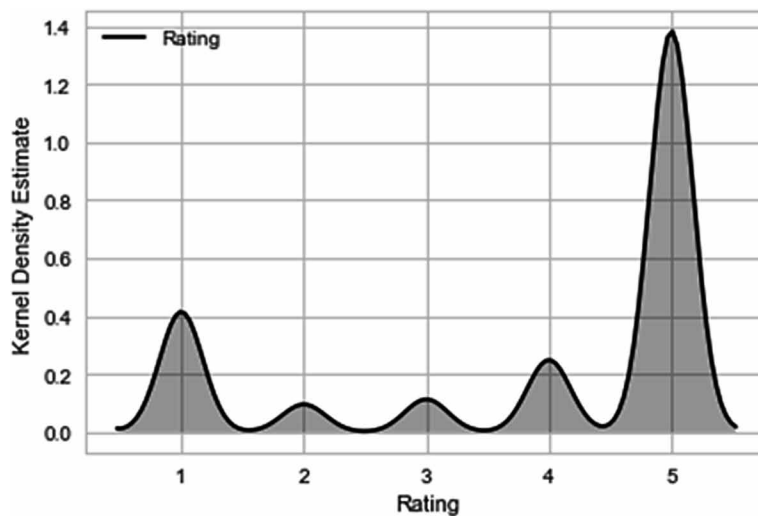


Figure 10. Kernel Density Estimate



to the other categories. This would suggest that most customers were reasonably happy with their purchases. Interestingly, a Rating of 1 had the second highest volume but as per the above there are significantly less Negative reviews which on the face of it suggests that those who gave a Rating of 1 their corresponding sentiment score may be higher and not directly related to the rating, this is for now a question to be explored further. Interestingly the observable trend for Rating scores of 5 and Positive reviews seem fairly aligned so further analysis on this is needed.

Figure 16 shows a small cohort of reviews with the corresponding Sentiment and Polarity scores. This gives the reader a very quick taste of output from the code showing how each piece of text is given a score. A very brief scan of the results shows that it 'seems' accurate, taking record [2] as a positive review i.e. 'Very Pleased' and conversely looking at record [5] with the word problems mentioned the sentiment is -0.3 which seems to fit. This is obviously not indicative of the total population of



Figure 11. Samsung Kernel Density Estimate

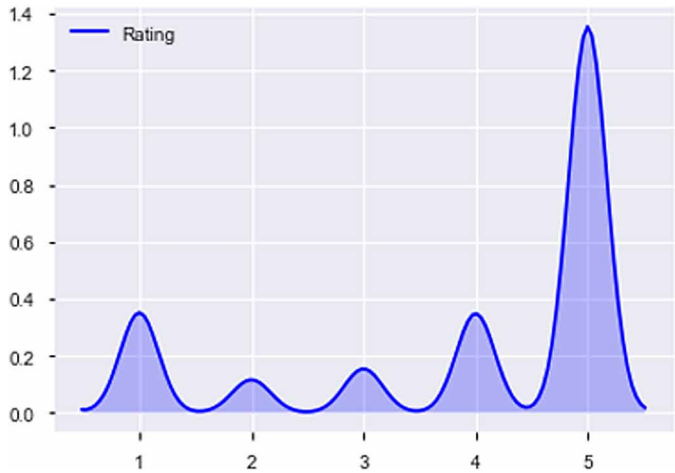
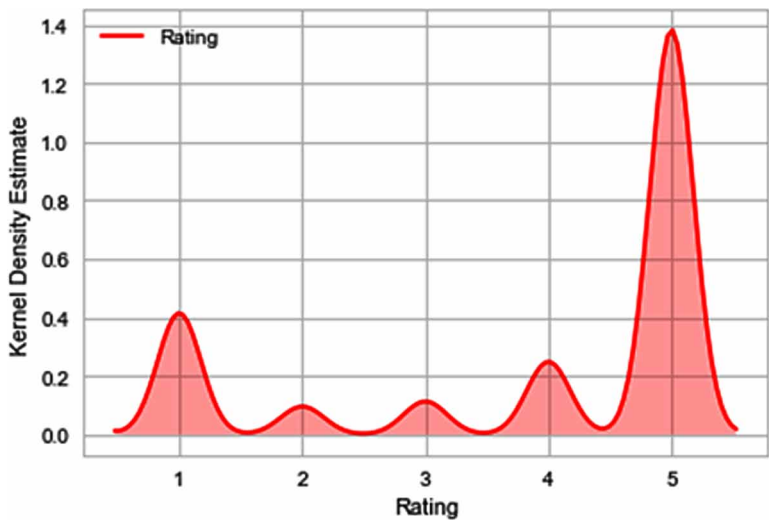


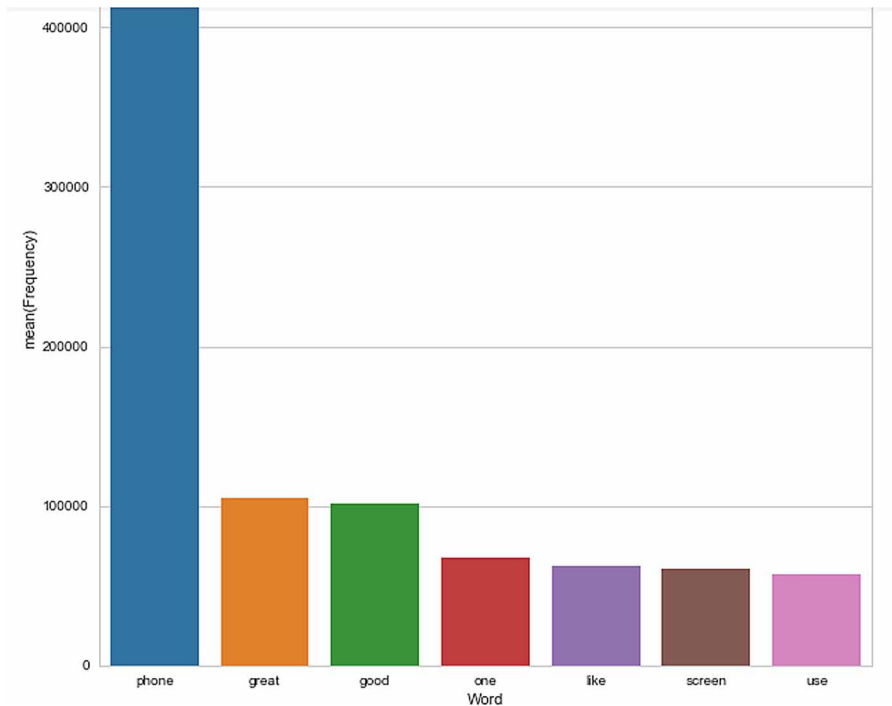
Figure 12. Apple Kernel Density Estimate



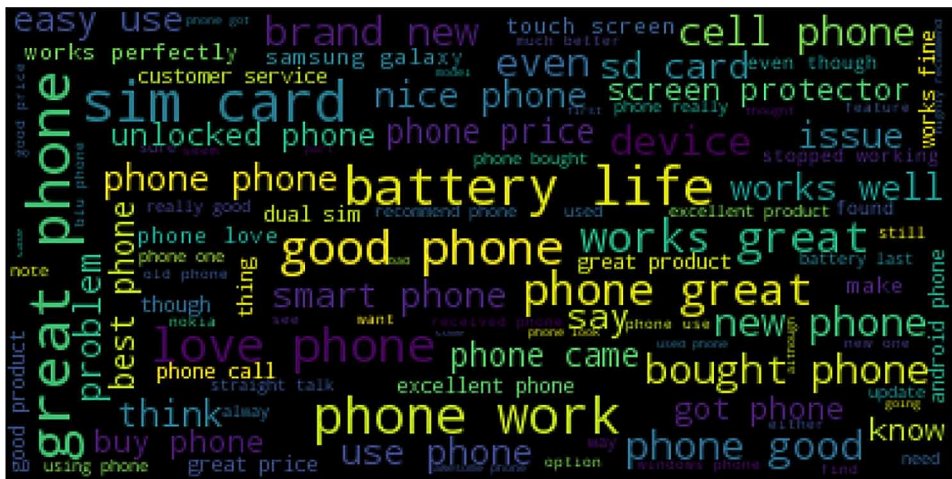
reviews but gives some insight into the results. Figure 17 shows the word cloud for Positive Reviews and here it is noticeable that key words such as ‘Great’, ‘Love’ and ‘Good’ are quite prevalent in the data and this is something that would be expected. Reviewing Figure 18, the word cloud for negative reviews it is noticeable that there are many occurrences of the words ‘Disappointed’, ‘Broken’ and ‘Bad’ which would tend to be on the negative side and again something that would be expected to see in the analysis.

The multinomial Naive Bayes classifier is suitable for classification with discrete features (e.g., word counts for text classification). The multinomial distribution normally requires integer feature counts. However, in practice, fractional counts such as tf-idf also work. The Multinomial Naïve Bayes model had a 93% accuracy rate when predicting a Rating of 1 and a 96% accuracy rate when predicting a Rating of 5, with an overall model accuracy of 96% The f1-score of 90% and 97%, with

### Figure 13. Word Frequency Chart



**Figure 14. Word Cloud**



an overall core of 96% represent the weighted average of the precision and recall scores indicating that the model is accurate when attempting to predict if a customer will score the review a 1 or 5.

The model was trained on the entire data set including all 5 ratings. Figure 19 shows the precision and recall output from the Multinomial Naïve Bayes modelling and Figure 20 shows the accuracy and averages. It is noticeable that when the model is run with the entire population of the data the accuracy drops quite significantly to an overall score of 74%. This indicates that the data for Ratings

Figure 15. Sentiment Frequency Chart

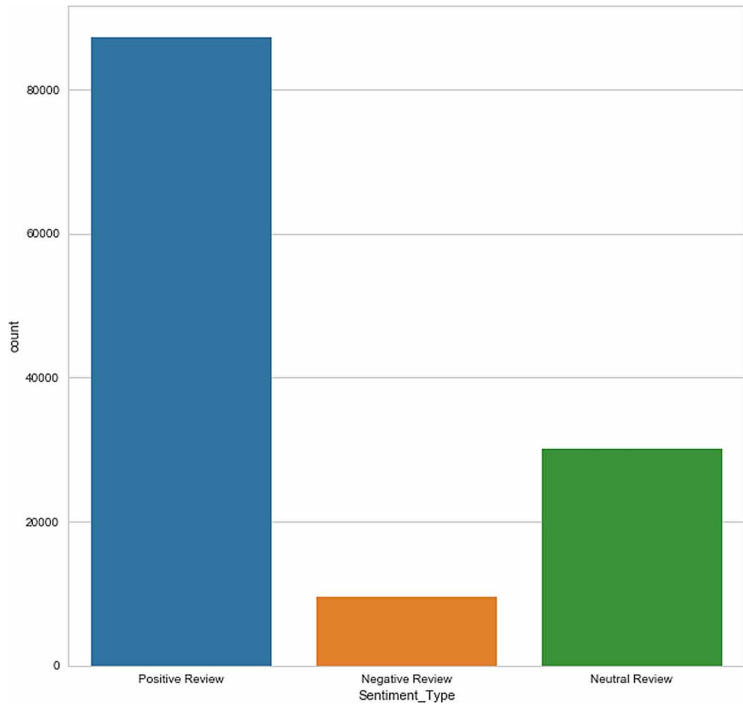
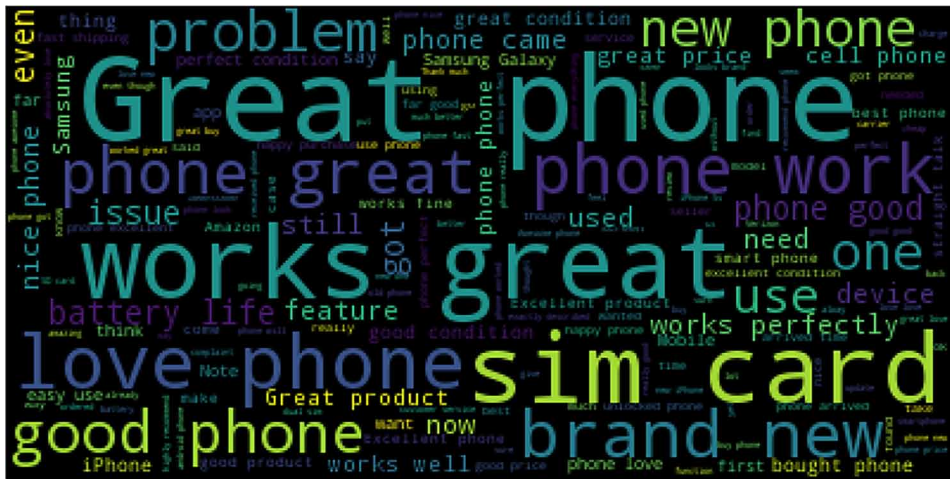


Figure 16. Reviews, Sentiment and Polarity Table

	Review	Polarity
0	I feel so LUCKY to have found this used (phone...	0.180952
1	nice phone, nice up grade from my pantach revu...	0.449259
2	Very pleased	0.650000
3	It works good but it goes slow sometimes but i...	0.452500
4	Great phone to replace my lost phone. The only...	0.214583
5	I already had a phone with problems... I know ...	-0.300000
6	The charging port was loose. I got that solder...	0.019814
7	Phone looks good but wouldn't stay charged, ha...	0.516460
8	I originally was using the Samsung S2 Galaxy f...	0.267692
9	It's battery life is great. It's very responsi...	0.222222
10	My fiance had this phone previously, but cause...	0.358333
11	This is a great product it came after two days...	0.182500
12	These guys are the best! I had a little situat...	0.282500
13	I'm really disappointed about my phone and ser...	-0.083333
14	Ordered this phone as a replacement for the sa...	0.191591
15	Had this phone before and loved it but was not...	0.236111
16	I was able to get the phone I previously owned...	0.416667
17	I brought this phone as a replacement for my d...	0.473611

### Figure 17. Positive Review Word Cloud



**Figure 18. Negative Review Word Cloud**



2, 3 & 4 has a significant impact on the models' accuracy. It is noteworthy that Ratings 1 and 5 remain quite high and this is potentially due to the large volumes of Ratings in these categories compared with the other categories.

## 4.2. Normalisation of Polarity

Using the mean and standard deviation of the Polarity\_Norm, each rating was matched with its corresponding normalised polarity score. Figure 21 is interesting for many reasons as it can be observed that a customer's rating of 5 will have approximately a 60% probability that the sentiment score will be 4. Secondly there is a noticeable distribution between 3 and 4 which is in and around the neutral sentiment category. This slight shift away from a polarity score of 5 suggests that while most of sentiment may be positive there is a significant cohort of reviews that have a slight negative or neutral tone which would indicate an area for further exploration from a business perspective.

Figure 19. Confusion Matrix (additional variables)

[[	5026	18	65	140	751]
[	876	176	34	190	388]
[	554	17	341	292	918]
[	249	7	32	1252	3283]
[	280	13	61	456	21196]]

	precision	recall	f1-score	support
1	0.72	0.84	0.77	6000
2	0.76	0.11	0.19	1664
3	0.64	0.16	0.26	2122
4	0.54	0.26	0.35	4823
5	0.80	0.96	0.87	22006
avg / total	0.74	0.76	0.72	36615

Figure 20. Accuracy and Weighted Average Table

Rating	Accuracy	Weighted Average
1	72%	77%
2	76%	19%
3	64%	26%
4	54%	35%
5	80%	87%
Avg/Total	74%	72%

The same method was applied when plotting the graph in Figure 22. Considering that these are the probability distributions for Ratings of 1, one could be forgiven for thinking these were the results for a Rating of 3. It is significant to observe that the polarity or sentiment of 1 Ratings are tending towards slightly negative and natural, this could be down to many reasons. The model itself may be unable to identify negative words, or customers tend to not be overly critical when commenting.

Figure 23 shows the distributions of Ratings 1,3,4 & 5 to give an idea of the how the sentiment has been measured for these ratings. Due to space constraints it was decided that Rating 2 would be left out and it also had the lowest volume of all the ratings. Rating 3 and 4 output graphs seem to be consistent and in line with the expectations in terms of the distribution of polarity. For that reason, the suggestion would be that no further analysis is needed for these scores and from a business perspective the most valuable insights come from the Ratings 1 and 5.

The model was intuitive enough to establish that sentiment analysis is a key indicator of the customer experience and for that reason should be seriously considered as part of any businesses metrics strategy. The key analysis and ultimately the findings of this research centre around the normalisation of the polarity score from the sentiment analysis to align with the corresponding rating given by the customer for that product. It was noticeable that a large distribution of ratings fell into



Figure 21. Normalised Polarity Distribution for Rating 5

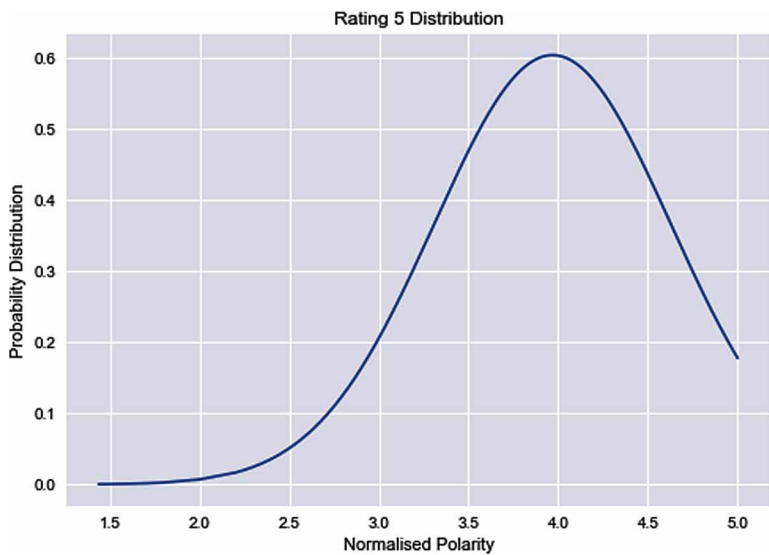
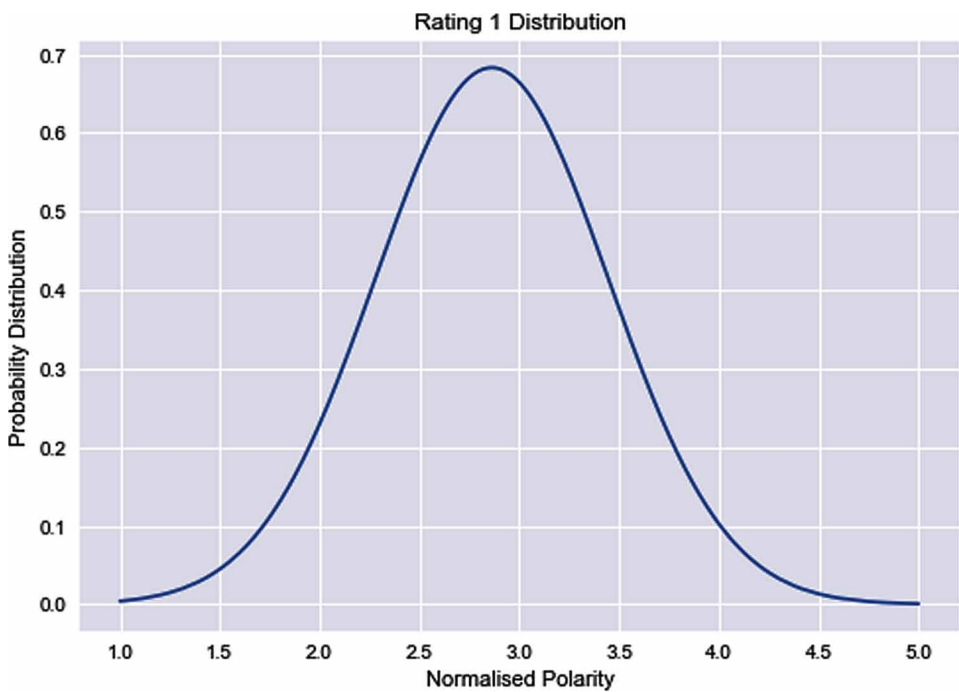
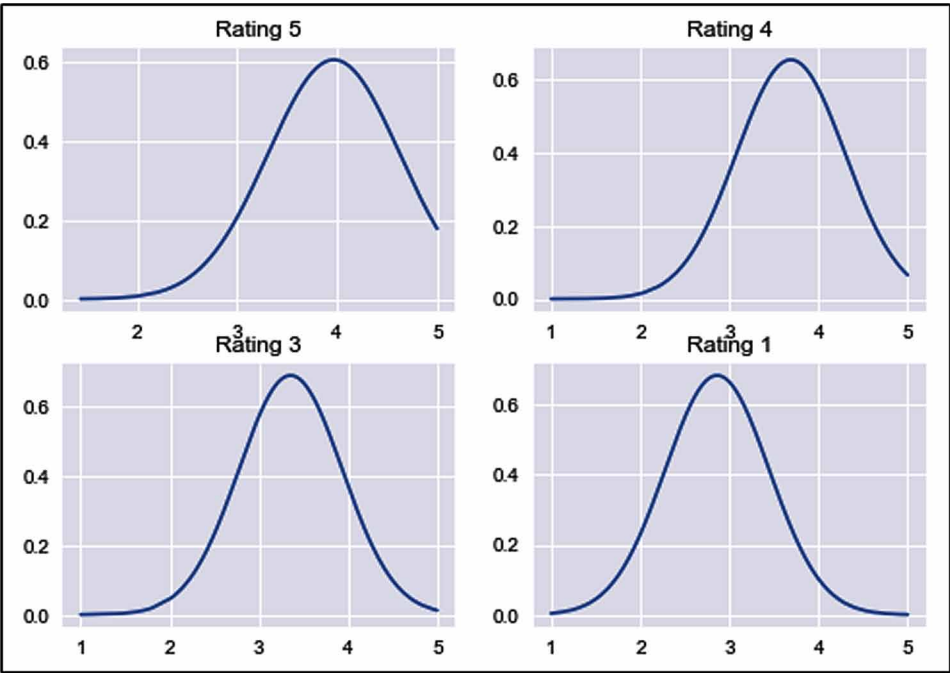


Figure 22. Normalised Polarity Distribution for Rating 1



the ‘1’ and ‘5’ buckets. For that reason, it was prudent to examine the associated polarity with these scores more closely than the other rating scores. The analysis, supplemented with the visualisations showed that for a Rating of 5, the sentiment distribution was skewed towards a score of 4 with a slight emphasis on 3 also. This showed that while customers were rating the products 5/5 their actual sentiment regarding the product was not so overwhelmingly positive. Conversely the same can be

Figure 23. Rating 1,3,4 & 5 Polarity Distribution



said of customers giving the products a rating of ‘1’. The results showed that while customers rated some products poorly the overall sentiment fell into the somewhat negative and neutral bands. What we show is that there can be an over reliance on using the Ratings score as a single point of truth for a business, when big data analytics techniques are employed it is possible to drill down into the real feelings of customers and get a clearer picture of products and services. The predictive analytics model used multinomial Bayesian networks to predict customer ratings using the sentiment review score. The results of the research highlight the short-comings of the ratings score as a single customer experience metric, thereby supporting the notion that a business cannot rely solely on this metric. The result from the prediction phase is a model that is capable of correctly predicting 74 per cent of customers ratings based on their reviews. This is not remarkably high percentage in terms of accuracy and backs up the claim above that the ratings cannot be used as an accurate predictor of sentiment alone. On this basis of the results, this research recommends that businesses are discouraged from using the CSAT or ratings scores as a single point of truth for customer experience. Although these metrics are important, they do not provide an insight into the opinion of customers regarding the experience they have had with a business or a product. The approach employed here demonstrates that a data-driven approach to exploring customer sentiment via text mining helps form a clearer picture for the business and it is important that firms capitalize on this type of data.

## 5. CONCLUSION

Quantitative survey questions such as Customer Satisfaction will always remain widely popular with businesses when attempting to take the pulse from their customers on products and customer experience (Hague & Hague, 2018). To overly rely on single customer metrics such as customer experience or net promoter score is a risky strategy and the suggestion is that business should focus upon a more nuanced multidimensional approach to predicting customer behaviour (Zaki et al., 2016).

We explored the notion of focussing and adopting a more scientific approach to measure customer sentiment by mining verbatim feedback and applying a sentiment analysis methodology to determine how customers feel about the products as opposed to the score they gave the product based on the constrictive 1 to 5 rating. We provided a framework to understand how a customer experience process can be developed to provide insights into the Voice of the Customer.

Key recommendations are as follows:

- Establishing and maintaining a positive customer experience for customers is imperative for any company's long-term goals and ultimately their success as a business.
- There needs to be buy-in at a senior management level to embrace a customer-centric metrics strategy which includes development of a robust process to measure customer sentiment.
- Text mining and sentiment analysis play a key role in understanding how customers feel about the products they purchase or the experience they had with a company (Fan, 2014).
- While survey-based metrics gathering will always remain an important process in measuring leading indicators of customers' behaviour, this should be supplemented with a formal customer experience programs that monitor performance and guide improvement efforts (Homburg et al., 2015).
- Organizations should begin to tap into the rich vein of customer feedback data at their disposal in the form of free verbatim text.

The central emphasis of the outcomes suggests that it would be folly for a business to ignore customer feedback from surveys or social media channels and solely rely upon quantitative such as customer satisfaction CSAT. There is no doubt that CSAT is an extremely important metric and should continue to be measured however businesses need to capitalise on customer data points such as sentiment to really get an understanding of how customers are feeling.

Descriptive analytics plays a significant role for a business in terms of understanding what is happening with their products and services. It gives a snapshot, usually after the fact of performance and experience and this research showed that by visualising these data points and reporting out on them can be valuable to a business. The results give a clear insight into customer opinion and the nuances around which customers think as opposed to how they rate a product on a limited 5-point scale. The findings point towards deeper learning around sentiment and how it can be utilised in other areas of the business and with additional data points not present in this research. This strategy could be employed in a business scenario and the results and findings would prove valuable in helping build a customer experience strategy for a business. The key findings are that it is unwise to rely solely on CSAT or NPS scores without examining customer feedback.



## REFERENCES

- ACSI. (2018). *American Customer Satisfaction Index*. Retrieved from <http://www.theacsi.org/about-acsi/the-science-of-customer-satisfaction>
- Ajenstat, F. (2017). Tableau five years a leader in Gartner's Magic Quadrant for Analytics. *Tableau*. Retrieved from <https://www.tableau.com/about/blog/2017/2/tableau-five-years-leader-gartners-magic-quadrant-analytics-66133>
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. arXiv:1707.02919
- Arokiasamy, A. (2013). The impact of customer satisfaction on customer loyalty. *Journal of Commerce*, 5(1), 14–21.
- Aswani, S. (2017). Analyzing Customer Feedback Data: Manual Analysis vs NLP. *Clarabridge*. Retrieved from <https://www.clarabridge.com/blog/analyzing-customer-feedback-data-manual-analysis-vs-nlp/>
- Baars, H., & Kemper, H. G. (2008). Management support with structured and unstructured data—an integrated business intelligence framework. *Information Systems Management*, 25(2), 132–148.
- Balahur, A. & Montoyo, A. (2008). A feature dependent method for opinion mining and classification.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52(2), 21–32.
- Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. In *Natural language processing with Python: analyzing text with the natural language toolkit* (pp. 1–34). O'Reilly Media Inc.
- Broan, S.E. & Steger, A.J. (2013). Key performance indicators are not just about profit. CFMA. Retrieved from <http://www.cfma.org/content.cfm?ItemNumber=1899>
- Brownlee, J. (2016). *Support Vector Machines for Machine Learning*. Retrieved from <https://machinelearningmastery.com/support-vector-machines-for-machine-learning/>
- Cappelli, A. (2017). *Natural Language Processing with Stanford CoreNLP*. Retrieved from <https://cloudacademy.com/blog/natural-language-processing-stanford-corenlp-2/>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Sherer, C., & Wirth, R. (2000). Cross industry standard process for data mining (CRISP-DM) 1.0.
- Chen, H., Chiang, R., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *Management Information Systems Quarterly*, 1(1), 1165–1188. doi:10.2307/41703503
- Chetviorkin, I. & Loukachevitch, N. (2013). Evaluating sentiment analysis systems in Russian.
- Chowdhury, G. (2005). Natural language processing. *Annual Review of Information Science & Technology*, 37(1), 51–89. doi:10.1002/aris.1440370103
- Clarabridge. (2015). Retrieved from [www.clarabridge.com](http://www.clarabridge.com)
- Cran Project. (2018). *CRAN Task View: Natural Language Processing*. Retrieved from <https://cran.r-project.org/web/views/NaturalLanguageProcessing.html>
- Daume, H. III, & Marcu, D. (2006). Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 26, 101–126. doi:10.1613/jair.1872
- Department of Industry, Australian Government. (2018). Understand your customers. [REMOVED HYPERLINK FIELD]Retrieved from <https://www.business.gov.au/info/plan-and-start/start-your-business/what-is-customer-service/understand-your-customers>
- Devika, M. D. C, S. & Ganesha, A., 2016. Sentiment Analysis:A Comparative Study On Different Approaches. Chennai, Tamil Nadu, India, Department of CSE, Vidya Academy of Science and Technology.
- EMC. (n.d.). [REMOVED HYPERLINK FIELD]Big Data storage [white paper]. Retrieved from <https://www.emc.com/collateral/white-papers/idg-bigdata-storage-wp.pdf>

- Schumaker, R. P., Zhang, Y., Huang, C. N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3), 458–464. doi:10.1016/j.dss.2012.03.001
- Fiedler, L., Großmaß, T., Roth, M., & Vetvik, O. J. (2016). [REMOVED HYPERLINK FIELD]Why customer analytics matter. *McKinsey*. Retrieved from <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/why-customer-analytics-matter>
- FoundationP. S. (2018). Get it. *Python*. Retrieved from <http://www.python.org/getit/>
- Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40(16), 6266–6282. doi:10.1016/j.eswa.2013.05.057
- Think with Google. (2016). Why customer analytics are the key to creating value. Retrieved from <https://www.thinkwithgoogle.com/intl/en-gb/marketing-resources/data-measurement/why-customer-analytics-are-the-key-to-creating-value/>
- Harris, D. (2018). What Is Text Analytics? We Analyze the Jargon. *Software advice*. Retrieved from <https://www.softwareadvice.com/resources/what-is-text-analytics/>
- Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2), 87–93. doi:10.2478/v10117-011-0021-1
- Hollyoake, M. (2009). The four pillars: Developing a 'bonded' business-to-business customer experience. *Journal of Database Marketing & Customer Strategy Management*, 16(2), 132–158. doi:10.1057/dbm.2009.14
- Hosseini, S., Maleki, A., & Gholamian, M. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, 37(7), 5259–5264. doi:10.1016/j.eswa.2009.12.070
- Huang, O. (2017). *Applying Multinomial Naive Bayes to NLP Problems: A Practical Explanation*. Medium. Retrieved from <https://medium.com/@Synced/applying-multinomial-naive-bayes-to-nlp-problems-a-practical-explanation-4f5271768ebf>
- Telus International. (2017). Customer First Magazine. Retrieved from <https://www.telusinternational.com/media/TI-CustomersFirstMagazine-I04.pdf>
- Kaisler, S. (2013). Big Data: Issues and Challenges Moving Forward. In *Proceedings of the 46th Hawaii International Conference on System Sciences* (pp. 995-1004). doi:10.1109/HICSS.2013.645
- Kim, E. (2017). *Everything You Wanted to Know about the Kernel Trick*. Retrieved from [http://www.eric-kim.net/eric-kim-net/posts/1/kernel\\_trick.html](http://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html)
- Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web* (pp. 342-351). ACM.
- Loria, S. (2018). *TextBlob*. Retrieved from <http://textblob.readthedocs.io/en/dev/>
- Maiolo, A. (2015). Comparing n-gram models. *Recognize Speech*. Retrieved from <http://recognize-speech.com/language-model/n-gram-model/comparison>
- Marbán, Ó., Mariscal, G., & Segovia, J. (2009). A data mining & knowledge discovery process model. In *Data mining and knowledge discovery in real life applications*. IntechOpen.
- Maritz. (2018). Maximise verbatims with text analytics [white paper]. Retrieved from <https://www.maritzcx.com/blog/wp-content/uploads/2014/11/Maritz-White-Paper-Maximise-verbatims-with-text-analytics.pdf>
- McCallumzy, A. K., & Nigamy, K. (1998, July). Employing EM and pool-based active learning for text classification. In *Proc. International Conference on Machine Learning (ICML)* (pp. 359-367).
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. doi:10.1016/j.asej.2014.04.011
- Mejova, Y. A. (2012). *Sentiment analysis within and across social media*. University of Iowa. doi:10.17077/etd.zze4b252

- Miner, G. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Oxford: Academic Press.
- Molag, T. (2017). Power BI vs Tableau. *Encore Business*. Retrieved from <https://www.encorebusiness.com/blog/power-bi-vs-tableau/>
- Natarajan, K., Li, J., & Koronios, A. (2010). Data mining techniques for data cleaning. In *Engineering Asset Lifecycle Management* (pp. 796–804). Springer London.
- NLTK.org. (2017). Retrieved from <http://www.nltk.org/>
- Ordenes, F., Theodoulidis, B., Burton, J., Gruber, T., & Zaki, M. (2014). Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research*, 17(3), 278–295. doi:10.1177/1094670514524625
- Ott, T. (2018). What is text analytics. *Software Advice*. Retrieved from <https://www.softwareadvice.com/resources/what-is-text-analytics/>
- Pang, B., & Lee, L. (2002). Thumbs up? Sentiment Classification using. *Machine Learning*.
- Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Understanding Customer Expectations of Service. *Sloan Management Review*, 32(3), 39.
- Project, C. (n.d.). *Cran r-project*. Retrieved from <https://cran.r-project.org/>
- Qualtrics. (2018). *What is the Net Promoter Score (NPS)*. Retrieved from <https://www.qualtrics.com/experience-management/customer/net-promoter-score/>
- Rapidminer. (2018). *Rapidminer*. Retrieved from <https://rapidminer.com/>
- ŘehůřekR. (2018). *gensim*. Retrieved from <https://radimrehurek.com/gensim/about.html>
- SAS. (2017). *Natural Language Processing*. Retrieved from [https://www.sas.com/en\\_us/insights/analytics/what-is-natural-language-processing-nlp.html](https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html)
- Smith, E. (2016). Everything you need to know about customer success. *Cobloom*. Retrieved from <https://www.cobloom.com/blog/everything-you-need-to-know-about-customer-success>
- Spacy.io. (n.d.). *Industrial-Strength NLP*. Retrieved from <https://spacy.io/>
- Srivastava, A., & Singh, D. M. (2014). Supervised SA of product reviews using Weighted k-NN Algorithm. In *2014 11th International Conference on Information Technology*.
- Srivastava, T. (2014). Support vector machine simplified. Retrieved from <https://www.analyticsvidhya.com/blog/2014/10/support-vector-machine-simplified/>
- Tamilselvi, A. & Parveen Taj, M. (2013). Sentiment Analysis of Micro blogs using Opinion Mining Classification Algorithm. *International Journal of Science and Research*, 2(10), 2319–7064.
- Taylor, C. (2018). [REMOVED HYPERLINK FIELD]Structured vs unstructured data. *Datamation*. Retrieved from <https://www.datamation.com/big-data/structured-vs-unstructured-data.html>
- Tirunillai, S., & Tellis, G. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *JMR, Journal of Marketing Research*, 51(4), 463–479. doi:10.1509/jmr.12.0106
- Upadhyay, P. (2017). Removing stop words NLTK python. *Geeksforgeeks.com*. Retrieved from <https://www.geeksforgeeks.org/removing-stop-words-nltk-python/>
- Verint. (2016). *Verint Text Analytics*. Retrieved from <https://www.verint.com/Assets/resources/resource-types/datasheets/text-analytics-datasheet.pdf>
- Vryniotis, V. (2013). Machine Learning Tutorial: The Max Entropy Text Classifier. *Datumbbox*. Retrieved from <http://blog.datumbbox.com/machine-learning-tutorial-the-max-entropy-text-classifier/>

Wichmann, M. (2017). *The Python Wiki*. Retrieved from <https://wiki.python.org/moin/>

Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*, 181(6), 1138–1152. doi:10.1016/j.ins.2010.11.023

Yadav, V., & Elchuri, H. (2013). Serendio: Simple and Practical lexicon based approach to Sentiment Analysis. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics* (pp. 543-548).

## ENDNOTE

<sup>1</sup> <https://www.kaggle.com>

*Kevin Curran is a Professor of Cyber Security, Executive Co-Director of the Legal innovation Centre and group leader of the Ambient Intelligence & Virtual Worlds Research Group at Ulster University. He is also a senior member of the IEEE. Prof. Curran is perhaps most well-known for his work on location positioning within indoor environments, blockchain and Internet security evidenced by over 800 published works. His expertise has been acknowledged by invitations to present his work at international conferences, overseas universities and research laboratories. He is a regular contributor on TV and radio and in trade and consumer IT magazines.*